AI and Design Opportunities for Smart Speakers

Tao Long, [1] **Lydia Chilton** [1]

[1]Computational Design Lab, Department of Computer Science, Columbia University
long@cs.columbia.edu, chilton@cs.columbia.edu

Abstract

Advances in voice technology and voice user interfaces (VUIs) — such as Alexa, Siri, and Google Home — have opened up the potential for many new types of interaction. However, despite the potential of these devices reflected by the growing market size and body of VUI research, there is a lingering sense that the technology is still underused. In this paper, we conducted a systematic literature review of 35 papers to identify and synthesize 123 VUI design guidelines into five themes. Additionally, we conducted semi-structured interviews with 15 VUI users to understand their use and non-use of the technology. From the interviews, we distill four design challenges that contribute the most to disuse. Based on their (non-)use, we identify seven opportunity spaces for designers and AI developers to explore such as focusing on information support while multitasking (cooking, driving, childcare, etc), incorporating users' mental models for VUIs, and integrating calm design principles.

Introduction

Advances in voice technology and voice user interfaces (VUIs) such as Alexa, Siri, and Google Home have opened up the potential for many new types of interaction. These technologies automate simple tasks and enable hands-free experiences, providing convenience and efficiency. Firstly, being operated by voice instead of vision or hands, these technologies have the potential to assist users during multitasking — when their hands are busy and attention is not fully focused — such as checking the temperature outside while dressing or changing the music while driving. Secondly, as they are situated in our home environment, they can help us with environment-related tasks like setting times in the kitchen while cooking, turning off all the lights before bed, and ordering stuff online while finding there is a shortage of something. Additionally, VUIs are often in shared spaces and can thus be used by multiple people at once — anyone can change the music in the living room without the need to hold a controlling device like a remote or a phone. In many ways, these devices have created the potential for the long-standing HCI visions of usable computing by enabling ubiquitous computing as well as accessible and calm design.

However, despite the potential of these devices, there is a lingering sense among customers that the technology is still underused. Both user experience studies (Mavrina et al. 2022; Trajkova and Martin-Hammond 2020; Voit et al. 2020; Kim 2021; Cho, Lee, and Lee 2019; Lopatovska et al. 2019) and marketplace review analysis (Fruchter and Liccardi 2018) report that VUI users are infrequently using or even abandoning the technology after their purchase. Multiple reasons contribute to the minimal use and non-use: limited use cases, cumbersome setup experiences, and difficulty in discovering new use cases (Trajkova and Martin-Hammond 2020). However, it is unclear how to design VUIs better such that their potential might be reached.

To understand the landscape of known challenges in designing VUIs, we conducted a systematic literature review of 35 research papers containing design principles. These papers either provide VUI design recommendations or evaluation heuristics. After a thematic analysis, we synthesized 123 design guidelines from these papers into five themes: 1) enhance basic usability, 2) customize for user contexts, 3) speak users' language, 4) design simpler interactions, and 5) establish reliability. Each theme contains multiple subthemes with actionable pathways to realize the theme and improve user experience.

To provide a concise list of AI design challenges and opportunities, we conducted in-depth interviews with 15 VUI users to explore and understand the reason for their use and non-use of the technology. We asked about their general experiences with their VUI device(s), with an emphasis on what was and was not useful about the device, including their daily interaction with the device, degrees of understanding of VUI-specific concepts, and future expectations. Analyzing the interview data using affinity diagramming, we found that there exist four major challenges that contribute the most to their non-use of the technology: 1) inefficiency of input and output, 2) gap between users' expected and VUIs' actual capabilities, 3) poor discoverability and 4) the lack of affective responsiveness.

Based on the successful use cases in the interviews, we present seven opportunities for designers and AI developers in building future VUIs: focusing on information support while multitasking (cooking, driving, showering, childcare, etc.), incorporating users' mental models for VUIs, and integrating calm design principles into AI design.

Related Works

History of VUIs and Smart Speakers

In the early 2000s, VUIs were introduced to the general public through interactive voice response systems which operated over the telephone. These systems could understand human speech and complete simple tasks (Pearl 2017). By 2011, voice recognition had improved to the point where Apple integrated Siri into mobile devices enabling hand-free interaction and popularizing the idea of a voice assistant (Gupta and Carew 2012; Allworth 2014). Steady improvements in machine learning and the availability of significant datasets fuelling them continually improved the quality of speech recognition and interaction with the technology (Team 2017; Reeves et al. 2018).

In 2014, the concept of smart home and modern home assistants appeared as Apple released its first home kit. Later, many leading technology companies like Google, Amazon, Xiao Mi, etc. also released their voice assistant specifically aim to control all smart devices in a household setting. Working actively with users in their daily life by automating simple tasks and enabling hands-free interactions, these new stand-alone devices opened up great potential for new interaction between the users and the voice technology. As these companies are vying for control of your home as well as the increasing stay-at-home behavior during the pandemic, more people adopt a device-based home assistant (National Public Radio 2020). In 2017, around 46% of American adults were using voice assistants while only 8% of Americans were using the technology via a separate stand-alone device like Alexa or Google Home (Olmstead 2017). In 2022, 62% of American adults are using voice assistants, and 35% of American adults are using their own stand-alone smart speaker devices at home (National Public Radio 2022).

Ubiquitous Computing and Calm Technology

Mark Weiser's vision of ubiquitous computing (Weiser 1991) depicted a future with computers and screens of many sizes embedded into all aspects of our lives. As a new way of thinking about technologies, these computers serve in concert to help people complete tasks without effort. With the wireless communication that ties all the diverse functions of the computers, Weiser believes that in the future homes, offices, campuses, and cities, there will contain hundreds of these tiny computers around us (Weiser 1991, 1998). Although VUIs were not the explicit focus of the original vision, it is still consistent with the notion of having quick and easy access to information under different circumstances.

However, this proliferation of technology comes with a serious downside. As users receive more information and have complex interactions with the devices, they experience sensory overload (Shedroff 2000). Experiencing multiple sensory cues (visual, audio, vibration, etc.) at the same time (Feng, Dey, and Lindeman 2016), receiving constant notifications (Pielot and Rello 2015), and having long-time online social interaction (Matthes et al. 2020) overwhelm our ability to process and respond to information.

Responding to information overload created by the high-maintenance technology, Mark Weiser and John Seeley Brown pointed out the potential of *calm technology*. The theory of calm technology states technology should help users focus on the things that are important to them rather than creating panic (Weiser and Brown 1995, 1996). Amber Case also published a list of actionable guidelines (Case 2015) for designers in designing calm technology. Most of the guidelines shared the themes of "requiring the smallest possible amount of attention," "informing and creating calm," and "making use of the periphery of attention." Thus, the calm technology concept reflects the vision of making computing situated and invisible as well as making users feel calm and focused.

An attention graph for a tea kettle (See Figure 1) reflects how technology seeks for and captures attention from users during the water boiling process: the attention is relatively high when the kettle is being set up but diminishes when the user walks away from the kettle, and eventually the kettle is forgotten. When the kettle shouts, all attention is drawn back to the kettle's state, and the user runs to pick it up (Case 2015). As more and more similar non-calm designs like this appear in users' daily technology experiences, the high maintenance will overload users' senses. Thus, instead of capturing users' full attention in a short period of time, many calm technology designs suggest that designers should make use of both the center and the periphery of our attention to minimize the burden (Weiser and Brown 1995, 1996; Case 2015).

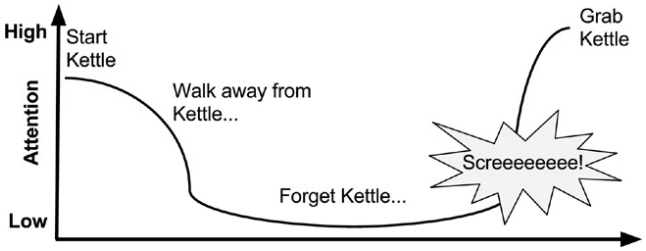

Figure 1: Attention Graph for a Boiling Tea Kettle by Amber Case (Case 2015). The graph shows how much attention is captured by a tea kettle while it's boiling water: the attention is relatively high when the kettle is being set up, but diminishes when the user walks away, and eventually the kettle is forgotten. When the kettle shouts, all attention is drawn back to the kettle, and the user runs to pick it up.

Ambient displays, serving as a screen that is always on without touching, provide an additional affordance to access more information and complete tasks while following the calm design vision by not dominating users' attention and cognitive load (Occhialini, van Essen, and Eggen 2011; Tentori, Segura, and Favela 2009; Kučera 2017; Jafarinaimi et al. 2005; Cho and Saakes 2017). Connecting with VUIs, this multimodal design helps visualize the information that is hard and inefficient to share verbally. Also, it contributes to the overall interaction experience as users could complete tasks quickly with more information given by the display (Yu et al. 2018).

Mental Models of VUIs

Recent research has shown that the mental model of technology informs people's perceptions of its role and capabilities (Gero et al. 2020; Davidoff et al. 2006). The strong personification of these smart speakers — humanized voices and human names — leads to users' belief that the devices can be modeled as human, thus making users intuitively refer to them as human and assume that it has human intelligence (Purington et al. 2017). Additionally, the term "smart" home and the devices' branding as "intelligent assistants" let people have high expectations of the system performance. This can lead to a mismatch between the expectations and the actual capabilities of VUIs (Mavrina et al. 2022). New users with little experience with VUIs tend to draw high expectations from their past experiences of human-human interaction (Luger and Sellen 2016; Cho, Lee, and Lee 2019). However, after finding the system lacks human-like abilities, new users often feel disappointed. Their mental model of the VUIs as having human-level intelligence and capabilities set them up for this disappointment.

Study 1: Systematic Literature Review

Through a systematic review of 35 related literature, we want to understand what themes are shared among the existing VUI design and development guidelines. In total, we identified and synthesized 123 guidelines from the papers that offer design recommendations or evaluation heuristics.

Method

We found 35 related literature that provides future design guidelines for a voice interface, intelligent voice assistant, or Alexa through the ACM Digital Library and Google Scholar. Using and combining a variety of terms to form our search queries, including "voice interface," "user experience," "design," and "Alexa," we later found an extensive list of articles. They cover a wide range of human-centered design research topics: designing for marginalized groups, collaboration work, household usage, and personal skill developments and training. 30 out of these 35 papers conducted user studies, through experiments, diary studies, focus groups, semi-structured interviews, etc. 10 of the papers conducted literature reviews to map future design guidelines, and 5 papers shared both methodologies. Also, more than half of the papers are published at ACM CHI Conference on Human Factors in Computing Systems (18 out of 35 papers) and in recent two years (20 out of 35 papers). (See Appendix A. Figure 2 for Detailed Descriptive Data Visualization).

Research Questions:

- What are some themes that the design guidelines shared across these papers?
- What are some higher-level visions and lower-level sub-themes behind the themes?

All 35 papers provide guidelines: 14 of them generate design recommendations, while 19 of them formulate evaluation heuristics. Finally, identifying and extracting all the design recommendations and evaluation heuristics, we found 123 VUI design guidelines in total. We put all the guidelines into a shared list and generated a basic codebook by reading through them. We analyzed and synthesized the guidelines to find the key similarities among their directions and visions reflected in their design guidance. The final themes and subthemes have been iterated over ten times to best capture the future vision given by the design guidelines.

Findings

From our systematic literature review, we synthesized and generated five major themes that are shared among the 123 design guidelines: 1) Enhance Basic Usability, 2) Customize for User Contexts, 3) Speak Users' Language, 4) Design for Simpler Interactions, and 5) Establish Reliability.

Under each theme, there are also a few sub-themes with a specific approach to realizing the theme. In total, we presented five major themes and fourteen sub-themes. The basic usability has six sub-themes, while the other themes only have two or one. (See Appendix B. Table 1. Lit Review Themes & Example Guidelines)

Theme 1. Enhance Basic Usability (52): Fifty-two guidelines mention enhancing VUIs' basic usability, which includes user control, error control, discoverability, multimodality, compatibility, etc. These major sub-themes all echoed Nielsen's heuristics (Jakob Nielsen 1994) and early evaluation heuristics for VUIs (Dybkjær and Bernsen 2001).

Promoting error prevention is mentioned in 21 guidelines. Users usually encounter unfamiliarity interacting with VUIs as it is so different from their previous experiences with other technologies (Begany, Sa, and Yuan 2015). Thus, designers should help users avoid the potential happening of errors by "making the system status clear (Murad et al. 2019)," "confirming input intelligently (Wei and Landay 2018)," and "handling ambiguous and underspecified utterance (Sciuto et al. 2018)." Providing VUIs more affordances to prevent errors, users could use the technology more confidently without encountering hardships during their early learning stage, thus prompting their adoption and frequent usage of the technology (Wei and Landay 2018).

Strengthening user control over VUIs is mentioned in 13 guidelines. Users could control and operate VUIs according to their preferences and needs: "enable users to share ownership with others (Zhang, Su, and Rekimoto 2022)," "mute or filter notifications (Vacher et al. 2015)," "provide abilities for users to control and interrupt (Maguire 2019)." Establishing a sense of control for users could help reduce errors and ambiguities while it is also key to generating security and trust in the system (Zhang, Su, and Rekimoto 2022).

Leveraging the multimodality of VUIs is mentioned in 7 guidelines. Incorporating visual, haptic, and audio interactions could provide users with more information while informing calmness. "Visually marking [the window that asks for attention] (Vacher et al. 2015)" and "making use of verbal and non-verbal cues (Axtell and Munteanu 2021)" do provide indication for understanding and enrich users' experiences with the system (Luo, Lee, and Choe 2020).

Improving system compatibility is mentioned in 4 guidelines. VUIs should be well integrated and function with users' frequently-use devices and services (López, Quesada,

and Guerrero 2018; Sciuto et al. 2018), like phones or Apps that people are using daily. Thus, the integration of the services with VUIs largely determines how much money users want to invest and how frequently users will interact with the technology ecosystem (Sciuto et al. 2018).

Enhancing system discoverability is mentioned in 4 guidelines. Low discoverability means that users cannot easily explore what features, interactions, and experiences are available to them. Due to a lack of a physical screen, VUIs could hardly share a new feature or related information quickly. Thus, leaving much potential by missing helpful interactions, users will just stay with the basic use cases and not learn new things (Sciuto et al. 2018). Thus, design guidelines mentioned easing the feature-finding process (Gollasch and Weber 2021), utilizing data mining (Sciuto et al. 2018), using previous responses (Striegl et al. 2021), or giving relevant feature recommendations (Kim 2021) to solve the low discoverability of the VUIs.

Facilitating error recovery is mentioned in 3 guidelines. Allowing users to easily "exit from errors or a mistaken conversation (Murad et al. 2019)" avoid users getting stuck inside an unsolvable problem. It makes users quickly learn the error and keep them in the natural workflow without frustration (Setlur and Tory 2022; Goetsu and Sakai 2019).

Theme 2. Customize using Users' Context (23): Integrating the usage history, user preferences, and users' background environment helps improve the interaction experiences, thus serving as a huge topic in future VUI design guidelines and heuristics. Responding to the mental model of VUIs, users may expect high efficiency and capabilities of communication from them to be "customizable", so designers should incorporate contexts to understand users' preferences and offer personalized interaction.

Remembering usage history and applying them to future automation is mentioned in 13 guidelines. Being different from the expectation, users found the VUIs' contextualization is still insufficient as they do not remember many important things as they expect AI "should be learning". Thus, many guidelines include providing specific accommodations for users' frequently-used commands (Pan et al. 2022). Also, based on users' needs, VUIs should make the interaction more efficient and useful: "Maintaining user profiles to delivering personalized experiences (Sabir, Lafontaine, and Das 2022)" through "associating contents and commands with specific users (Zhang, Su, and Rekimoto 2022)" or "employing territorial markers to help them avoid activity- and preference-related conflicts (Zhang, Su, and Rekimoto 2022)", the context-enabled automation will help users' interactions with VUIs become more efficient and effective (Pan et al. 2022; Axtell and Munteanu 2021).

Designing user-centered proactive adaption for special user groups, like parent-kids interactions, seniors, and people with diverse abilities, is mentioned in 6 guidelines. For example, the amount of the VUI automation involved in the interaction (Zhang et al. 2022) and responding contents (Kim 2021) should be adapted based on the capabilities and background of the users. Also, designers should utilize the "social-justice oriented design" and spend more effort understanding tools delegated for "the overburdened and under-appreciated workforce. (Bartle et al. 2022)"

Leveraging the physical background environment is mentioned in 4 guidelines. Similar to Mark Weiser's idea on ubiquitous computing, many guidelines suggest using sensors (Reddy et al. 2021) and location markers to "modify its level of proactivity, listen for particular commands, and offer more appropriate suggestions for users (Zhang et al. 2022)."

Theme 3. Speak Users' language (23): Many guidelines mentioned that the VUIs should speak users' language including using conversation-like language and expressing human emotions throughout the interaction. (Yang, Aurisicchio, and Baxter 2019). This topic echoes the users' mental model of smart assistants as they are expecting high adaptability of the system to their needs and preferences. Filling the gap in (Murad et al. 2018), this theme also relates to the consistency and naturalness in Nielsen and other heuristics (Jakob Nielsen 1994; Dybkjær and Bernsen 2001).

Accommodating conversational language is mentioned in 17 guidelines. Future designers should make VUIs use human-like language and patterns to sound more natural (Kim et al. 2021). These human-like conversation "tricks" includes turn-taking (Alrumayh, Lehman, and Tan 2020; Wei and Landay 2018; Murad et al. 2021), re-mentioning previous contents in the chat history (Wang et al. 2020), back-channeling (Cho et al. 2022), and understanding and applying social cues (Alrumayh, Lehman, and Tan 2020). Human-like conversations and communication strategies will build trust between the users and the system, thus leading to frequent usages (Davidoff et al. 2006).

Anthropomorphizing human emotions and expressing affective responsiveness are mentioned in 6 guidelines. Designers should make sure VUIs could express interest to users: show empathy and emotional responsiveness (Kim et al. 2021; Wang et al. 2020). As the interaction between users and VUIs becomes more affective and natural, users will become more engaged and establish more trust in the system (Kim et al. 2021).

Theme 4. Design Simpler Interaction (16): Many design recommendations and evaluation heuristics focus on the conciseness and simplicity of the interactions between users and VUIs. Responding to Mark Weiser's vision (Weiser and Brown 1995, 1996), the conciseness and the simplicity both help users minimize cognitive load (Alrumayh, Lehman, and Tan 2020) and optimize the users' time, thus allowing users to focus on the information that is more crucial (Nowacki, Gordeeva, and Lizé 2020). Filling the gap in (Murad et al. 2018), it also echoes both the "minimalist design" in Nielsen's heuristics (Jakob Nielsen 1994) and the "the right amount of technology is the minimum needed to solve the problem" in *Calm Technology* (Case 2015).

Designing for shorter conversation is mentioned in 8 guidelines. Keeping the input and output concise as well as acknowledging users before a long interaction are important (Wei and Landay 2018). Thus, guidelines mention "maximize efficiency (Zwakman, Pal, and Arpnikanondt 2021)" and "be concise and to the point based on the user's intent (Alrumayh, Lehman, and Tan 2020),"

Designing for simple conversation between users and VUIs is mentioned in 8 other guidelines. In addition to time efficiency, VUI designers should keep the interaction structurally (Wei and Landay 2018) and acoustically (Zwakman, Pal, and Arpnikanondt 2021) simple to avoid confusion.

Theme 5. Establish Reliability (9): As a huge section in security and privacy, the reliability of VUIs is largely concerned. People are always concerned the technology is always listening and stealing personal data as they are taking advantage of its special setting in users' homes (Lau, Zimmerman, and Schaub 2016; Malkin, Egelman, and Wagner 2019; Tabassum et al. 2019; Javed, Sethi, and Jadoun 2019).

Establishing a reliable perception of VUIs is mentioned in 9 guidelines. It includes "providing an additional feedback (Reddy et al. 2021; Alrumayh, Lehman, and Tan 2020)," explicitly including warning messages and consequences (Zubatiy et al. 2021; Chalhoub et al. 2021), and indicating system contexts between native and third-party skills for the users (Major et al. 2021).

Study 2: User Interviews

To better understand users' challenges that lead to their non-use, we conducted 15 in-depth interviews with Alexa users. From there, we distilled insights about their obstacles in using Alexa and generated future design opportunities to bridge the challenges of non-use.

Method

Between October 2021 and February 2022, we conducted semi-structured interviews with 15 users of Alexa, who actively use Alexa in their daily life. We specifically chose Alexa to formulate our interview because it now shares the largest market share and it has been the leading brand of VUIs and smart speakers (Reportlinker 2022).

We reached out and recruited the participants through a college-wide email network that enabled connections with people from various backgrounds and age groups. Each interview session lasted for one hour. With exemption from the Institutional Review Board (IRB), we paid each participant $20 for their time. We anonymized all the names and identifiable information of the participants. Participants also received a consent form before their interview. With their consent, the interviews were conducted online through Zoom where we audio-recorded and transcribed the interviews. The user interviews aimed to understand the users' experiences with Alexa, with an emphasis on what was and was not useful or enjoyable about the device. We broke the interview process into sections to help us better understand their daily routine with Alexa, their enjoyment and frustration with Alexa, degrees of understanding of Alexa and Alexa-specific concepts like skills, and future expectations.

After conducting the interview, two individuals coded the interviews and picked the standing-out quotes. Then, they generated codes to find the overarching themes of the challenges and enjoyable moments they encountered during their usage. Finally, they list all the information on an affinity diagram that contains all the themes shared among users.

Findings: Challenges

From the user study, we found four major challenges that specifically contribute to users' non-use: 1) inefficiency of input and output, 2) the gap between users' expected and VUIs' actual capabilities, 3) poor discoverability and 4) lack of affective responsiveness. Mapping from these challenges that lead to non-use as well as their enjoyment with VUIs, we generated seven future opportunities for AI designers.

Challenge 1: Inefficiently of Input and Output Many users reported that they initially want to use Alexa since they want to finish the task as quickly as possible and easily since it provides a hand-free usage scenario. These save them time and effort in finding and using the traditional physical screens or objects to complete the task:

> *"Reducing the number of steps to get things... [has] been a trend in technology. The steps to complete a task are lower and lower until you don't have to do anything by hand."* (P9)

Saving users time and effort becomes the main reason and the driving expectation for the users to use Alexa to complete their tasks. However, opposite to the users' imagined capabilities of the VUIs, many users responded that they met many hardships coping with Alexa to complete a simple task efficiently, including the Alexa speaks so slowly (P10), it needs multiple rounds of back-and-forth conversations to complete a simple task (P1), Alexa cannot understand some complex queries that include many elements (P3), or they think Alexa always give many useless follow-up questions or hints that waste their time (P11, P14). P11 mentioned that Alexa always prompts to set up some irrelevant features after every command. For example, when they asked Alexa for the weather and temperature outside, Alexa would later ask them if they would like to set up another feature.

> *"[I thought] Alexa is an efficient speaker and alarm system... I always felt it slowed me down... I don't get it. I feel I can do anything in seconds."* (P5)

Thus, the contrast between the imagined efficiency and the actual tedious and irrelevant interactions contributes to user frustration. Users will quickly find alternative ways to complete the tasks and finally reach non-use.

Challenge 2: Gap between Users' Expected and VUIs' Actual Capabilities Similar to the highly-expected efficiency, users also overestimate the capabilities of Alexa in understanding queries and handling searches and answer-finding. Some users compare the search performance and experiences of Alexa devices to Google and other search engines (P1). However, Alexa sometimes understands users incorrectly and can hardly recover from errors.

> *"[I] wouldn't trust her to do a very complex search. She would misunderstand something along the way. She would interpret something else, or would get the words wrong... There's no point to easily go back. There's no "let me go back and clarify."* (P3)

Thus, after getting stuck in repeating the conversation all over or finding another way in completing the task, users

will get disappointed with the technology, which leads to non-use. However, due to the limitations and biases of Alexa's datasets, users often encounter hardship in completing tasks that involve special terms, thus leading to non-use.

Many users also discussed their problems with Alexa's recognizability in the unintentional waking. The mistake creates difficulties and unexpected experiences as Alexa interrupts and slows users' tasks.

> *"Whenever someone says Alexa/or something that sounds like it and she lights up and starts talking... mistaken interaction...Issues with talking to one and then the other gets activated."* (P12)

Alexa's recognizability also falls while taking queries that are culture-specific or from a user with an accent. Alexa has trouble understanding specific cultural or social terms that are tied to underrepresented cultural groups. For example, P7 mentions that Alexa could not understand them while communicating non-English song names, like Bollywood music, Latinx, and Afrobeats. Similarly, P15 shared that Alexa always messed up with Arabic music names so they have to manually select them. Also, P15 shared that they found their parents need to repeat themselves while using Alexa as Alexa hardly understands their Indian accent. Thus, most of the users who encountered hardships in recognizability ended up feeling disappointed at Alexa and using their phones to play music instead, thus leading to non-use.

Challenge 3: Poor Discoverability Many users also shared their challenges in learning and discovering new use cases in the technology. They met many troubles in adopting the device and integrating them closely into daily life. Firstly, users faced many difficulties during the system setup stage because there isn't enough information or previous experience in setting up a non-digital interface. For example, users were confused about the compatibility of Alexa while connecting them to other devices that they typically use. Also, due to the loss of visual screens, users could hardly explore the new use cases visually. Thus, finding it out of their habit and too time-consuming, users lost their motivation in finding new features or skills to use in Alexa.

> *"There's not enough information when you set up the device...[Everytime when they want to discover more features,] it was more than one step and I didn't have time to figure that out.."* (P1)

Due to the complexity of Alexa skills and related terminologies, users also face hardships in understanding how to use Alexa with the correct syntax. P3 and P8 mentioned that they learned and adjusted their query sentences after finding Alexa didn't get their message in the first place.

> *"[It] still feels like [you are] operating a command line where you have to use syntax. You're making it behave like a computer to speak to it. The whole point was conversational and natural. Let it be unhinged, closer to messy humans."* (P8)

Thus, hindering users' discovery of new features, users could hardly find new use cases to improve their usage of the technology. Without continuous and new interactions, users won't formulate a mental habit of using Alexa in replacing their mobile phones or laptops, which they typically have more experience and spend more time on. Thus, some users will use other technologies that have similar affordances.

> *"I'm a very visual person... For music, it feels easier on my phone: like the experience of scrolling through the playlists on my phone."* (P5)

Without a chance to learn the features closely, users like P5 may easily miss the chance of earning a mental model of the technology usage. Thus, the low discoverability and barriers in adopting a mental habit of using Alexa continuously make users feel the experiences out of habit. Users may experience minimal usage and eventually abandon the technology.

Challenge 4: Lack of Affective Responsiveness Many systems and user experience design problems on contextual relevance contribute to the challenges for users to accept Alexa. Even though literature shows that the personification of Alexa affectively connects users and the device (Purington et al. 2017), Alexa actually cannot understand users' moods nor express their own emotions with the users during the interactions.

> *"[VUIs have] no understanding of what mood you're in. It feels like a tool rather than a being. That's why it's hard to emotionally connect. You can't have an emotional or adult conversation with it. Although it cosplays as that. Feels like you are using a computer rather than talking to a person."* (P8)

P8 expressed disappointment about the VUIs' lack of emotions and they mentioned that they no longer express emotions in front of Alexa. As the affective connectivity between the users and VUIs contribute to the trust establishment and user engagement, the lack of affective responsiveness of the system will increase the risk of user non-use.

Findings: Opportunities

Responding to the challenge categories, we develop seven related opportunities for future AI designers and developers. Exploring further design directions, we aim to bridge the gap between users' needs or expectations and the future AI development of the smart speakers and their ecosystem.

Opportunity 1. Explore Use Cases to Understand and Support *Information* While *Primary Task* Many users reported that they use the device while doing some other tasks that partially or fully take their hands and attention away: situational impairments. For example, many usage scenarios are like playing music or radio while getting up (P15), asking for the weather while dressing (P12), hearing the news while making breakfast (P3), or setting timers while cooking (P13). If the users want to learn some other information from the VUIs while multitasking, they are definitely in need of a system based on voice interaction so that doesn't need to use their hands to touch the phone screens.

In addition to being hands-free, users also look for eye-free technology while dealing with some more complex and high-stake scenarios. For example, when parents want to ask about the weather outside while they are dressing their children, they could hardly pull their attention and hands away

from their children. Thus, users with similar special needs like parents, drivers, or the vision impaired, are looking for extra affordances and personalization of VUIs to accommodate this vision-free technology. ***How can AI Help?*** Powering users under these scenarios, a more targeted contextual information gathering and an in-depth segment analysis should be applied to the users' general usage scenarios as well as personalized ones according to the user's daily routine. After understanding users' persona better through more contextual signals and deep learning, special modes supporting personalization, understand an incomplete query, or enable shortcuts will be automatically turned on to fit users' needs and their situational impairment accordingly.

Opportunity 2. Focus on Short Commands rather than Extended Conversations Users find it difficult to have an extended conversation with VUIs. Instead, many successful interactions used single short commands to complete the tasks (P1, P5, P6, P8, P9, P14). For example, commands such as "Turn on the light", "Play Disney Classics on Spotify", and "Is ficus dangerous for dogs" are quick and achieve an important goal. This is consistent with the "one-breath test" (Blankenburg 2018) for Alexa developers which states that both questions and responses should be short enough that people can speak them in one breath. However, since people now are having more and more applications and devices, some of their commands will become more complicated. Thus, VUIs should be capable of taking and understanding a short but more complex command: a single-sentence command with multiple equally important contents. For example, the user wants to ask Alexa to play a certain song by their preferred singer from one of the streaming services, or the user wants to set three consecutive times in 5 min, 10 min, and 20 min to prepare for dinner. Thus, there will be at least three important elements in each single-sentence query. Currently, many VUIs still miss some of the components and give wrong results back to the users, thus leading to disappointment and non-use. ***How Can AI Help?*** Along with the advances in natural language processing especially in Question Answering and the contextual data generated by the users, VUIs could better capture the users intends from either short, long, and complicated commands.

Opportunity 3. Integration with Other Devices and Services Better integration and compatibility with various devices and App services matter to both users and developers. P7 mentioned the biggest turn-off is that Alexa does not integrate well with an iPhone while all Alexa users they know have iPhones. Currently, the hardship in developing VUIs, like Alexa, are largely resulting from ground-level system design, like the Alexa Presentation Language. For example, users could hardly search for and find their scheduled events stored in their calendars from their Alexa devices. In the first place, the developers face troubles in building more skills and connecting to more Apps or Web services that users tend to use, thus leading to the users' bad experiences in finding their wanted features on VUIs. According to much literature and heuristics on the compatibility and connected services, it is important to make the whole VUI system reachable for both users and developers to connect with other devices and platforms, and services. ***How Can AI Help?*** In addition to the development support for developers, better integrations with wearable devices helps gather more data from the users implicitly. Deep learning based on the diverse data inputs could power the personalization experiences and contribute to the ambient vision of VUIs.

Opportunity 4. Understand Emotions and Provide Affective Responsiveness The understanding of emotions could benefit the people who are frequently talking to the device and looking forward to emotional feedback. However, P8 shared that VUIs still lack emotional and affective responsiveness to their users. Previous literature (Wang et al. 2020) on using voice interfaces to help with communication and public speaking shows that users could receive validation and comfort from a voice agent, thus helping with establishing confidence and expressing themselves. ***How Can AI Help?*** Deep learning based on more frequent detections on the users' talking speed, speech tone, mood, and general usages of the verbal and nonverbal cues could help understand users' sentiment and affective experiences. With an intervention of emotions and affective voice responses adapting to users' needs, VUIs could establish a close relationship with users and replace human caregivers as an "integral member of people (Ramadan, Farah, and El Essrawi 2021)", thus prompting their future usage and future research understanding the affective aspects in human-agent interaction.

Opportunity 5. Smarter Interactions by Learning User Preferences There is a need to make the interaction shorter and smarter to fit into the quick pace of the user's life: from both literature reviews and user interviews, many examples show that people wanted to have more targeted and personalized responses (P8). For example, when it is going to rain outside, the VUIs should respond to the users who asked about the weather quickly the time when it will start raining, and how long it will last. In this case, VUIs should prioritize the contents that are delivered to the users based on users' preferences and the physical context. Here is the VUI's response after asking about the weather: "Currently, in [city name, state/country name], it is xxx degrees and light showers. Today, there will be heavy rain with a forecast high in xxx and a low in xxx. Due to the current humidity, it feels like it's xxxx degree." As the response gets to over three sentences and doesn't provide the exact start time of the rain, users could easily miss the point that they need to take an umbrella with them. However, it will also be helpful if the users could choose the information they want to hear or enable the "smart learning feature" to better understand their preferences. Thus, offering users the information they want and need to know under the context and their preferences is a plus for user experiences. ***How Can AI Help?*** Incorporating interactive machine learning, VUI systems could implicitly understand and capture users' intents to optimize experiences: automate the same tasks being used everyday, offering personalized shortcut, and generating supplemental information with multimodality.

Opportunity 6. Promote Social Learning from Observations and Tutorial Videos Discoverability of the device

sometimes largely relies on the company itself in the system design. However, external efforts could also contribute to this and help people learn new information about their device, thus possibly changing the users' non-use. In most cases, it's hard to learn about new tricks or hacks in using a smart home device since most of the usage is during private times and in their personal space. However, P2 mentions that they adopt many Alexa tricks from seeing Ad videos of Alexa: "There is a guy in the kitchen showing all the things he was doing with his Alexa while cooking. That's how I learned." Thus, as high-capacity media, videos and video-based social media platforms like TikTok or YouTube provide many tricks in using VUIs. For example, the viral tag #alexahack on TikTok has received 33.4 million views by Sep 2022. Thus, while seeing some openly shared tutorials online, users will quickly turn down their phones and interact with their devices to play with the new trick. Similarly, the social networks and groups run by VUI users like Reddit, Stackoverflow, or brand-specific forums like Alexa Developer forums are all powerful in sharing usage experiences, problems, and tricks of VUIs. ***How Can AI Help?*** With the advances of large language models and text-to-image generative models, users could easily create a sharable experiences with others. Thus, in this way, people who are interested in replicating online tutorials or videos could easily find one and learn from it, helping with their future convenience and use of the VUIs.

Opportunity 7. Build Towards a Future of Calm Technology An interesting future direction is to integrate more principles or considerations from calm design into the development and design of the VUI device. Similar to the calm technology and ubiquitous computing vision, future interactions with VUIs are considered to be efficient and focused. Users could use this assistant technology and integrate them well into daily life. P7 and P10 mentioned that they usually use VUIs to avoid looking at a digital screen or using their phone. However, technology companies use the metrics like users' screen-tapping frequency and screen time to evaluate user engagement. Thus, attention-seeking design and complicated interactions are built to ask users to stay on the screen and the APP, thus taking a lot of users' attention. In contrast, without the demand for much of your attention as the phone and efforts, VUI could complete users' tasks verbally, meeting their needs and providing a better experience. Previous works in the ambient display along with the VUIs also contribute to the promising future of the technology. As the ambient display requires little user attention but is more capable of delivering visual information, it will largely increase the information capacity that users could gather from the interactions in the VUI ecosystem. The development of ambient displays also opens up possibilities for exploring calm design principles both visually and verbally and future research on human-VUI interaction. ***How Can AI Help?*** Various devices powered by ubiquitous computing help gather users' data implicitly and easily. Without explicit inquries, the device could understand user profiles and intents progressively with interactive machine learning, thus delivering a more personalized and "calm" experience.

Limitations

Study 1: The list of the 35 papers takes over one month to finalize. Thus, even though we searched the database with the same queries, the search results varied throughout the month. Also, we only the default Relevance filter on ACM Digital Library and Google Scholar, so there must be some new papers that we missed from the search. Study 2: Participants are only Alexa users and they don't all have the same version of the Alexa device installed. Also, some participants have multiple devices and share devices with family members or roommates while others use one device individually. Also, we didn't interview any seniors or special populations who use Alexa or other VUIs, and we only focus on the users' perspectives. There are definitely many challenges and opportunities from the perspective of developers and stakeholders. We also limit the size of user interviews and literature reviews on guidelines.

Conclusion

To understand the design challenges that lead to non-use and to map future opportunities, this paper conducted a systematic literature review on 123 design guidelines from 35 publications in designing VUIs and user interviews with 15 users. Findings from the literature review present five major themes: 1) enhance basic usability, 2) customize for user contexts, 3) speak users' language, 4) design simpler interactions, and 5) establish reliability. From the interview, four challenges are identified in contributing to users' non-use: 1) inefficiency of input and output, 2) gap between users' expected and VUIs' actual capabilities, 3) poor discoverability and 4) the lack of affective responsiveness. Lastly, based on the insights and challenges, we mapped the future design opportunities for VUIs to improve user experiences and avoid user non-use, including focusing on information support while multitasking (cooking, driving, showering, childcare, etc), incorporating users' mental models for VUIs, and integrating calm design principles. Our work maps further vision of ubiquitous computing, calm design, and ambient AI technology along with the concept of smart home.

Appendix A

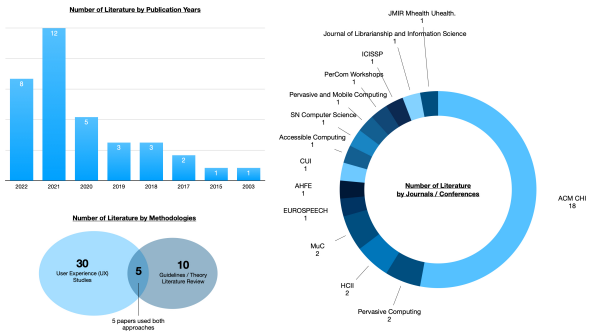

Figure 2: Descriptive Data for the 35 Literature

Appendix B

Table 1: Lit Review Themes and Example Guidelines

Note: The number (#) inside the parentheses indicates how many guidelines belong to the theme.

Themes	Sub-themes	Example Guidelines
Enhance Basic Usability (52)	Promote Error Prevention (21)	A12. Confirm input intelligently: [current VUIs] failed to explicitly confirm some critical actions [, like double-checking which alarm to turn off] (Wei and Landay 2018).
		[VUIs] need to provide feedback to the user explaining their interpretation of the [ambiguous or underspecified utterance] and how it was handled (Setlur and Tory 2022).
	Strengthen User Control (13)	Provide ability for users to control and interrupt (Vacher et al. 2015).
		The control users have on the processing of their actions by the system. (Nowacki, Gordeeva, and Lizé 2020).
	Leverage Multimodality (7)	Address [low situation awareness] by providing better vocal feedback... and adding visual indication for information that is critical to the user (Luria, Hoffman, and Zuckerman 2017)
		Leverage task context and multimodality in order to provide visual or other non-verbal cues (Axtell and Munteanu 2021).
	Improve Compatibility (4)	Smart Home Framework: the compatibility of the [VUIs] with smart home devices (López, Quesada, and Guerrero 2018).
		Integrate [the VUIs with] not only smartphones but also connected televisions, computers, and other screen-based devices (Sciuto et al. 2018).
	Enhance Discoverability (4)	Use responses to help users discover what is possible...rather than always say something is impossible: the system did not teach ways to ask for a result, and [users] had to guess and try multiple times (Wei and Landay 2018).
		Data mining to offer new features: ...significant challenge for [VUIs]. One opportunity is to data mine repeated patterns of use or use common routines as scaffolding to introduce new related features (Sciuto et al. 2018).
	Facilitate Error Recovery (3)	Provide interface affordances (visual or language) so users can refine and repair system choices (Setlur and Tory 2022).
		A17. Allow users to exit from errors or a mistaken conversation: Use a special escape word globally (e.g. "Stop")... [or] non-speech methods when speech fails (e.g., push a physical button) (Wei and Landay 2018).
Customize for User Contexts (23)	Remember Usage History (13)	Remember User Profiles to Deliver Personalized Services: store a vast amount of information specific to the user, such as personal profiles or preferences (Kim et al. 2021)
		Enable user to employ territorial markers to help them avoid activity- and preference-related conflicts by communicating their preferences or staking a claim to the device or data (Garg 2022).
	Design for Diverse & Sensitive Populations (6)	Enriching the responding contents when executing these commands might give older adults a chance to find more features and functionalities (Kim 2021).
		Adopt social-justice oriented design methods...when building [VUIs] in home health care contexts (Bartle et al. 2022).
	Leverage User's Environment (4)	Integrate the functionality of ubicomp sensors: install additional sensors in the home that can serve as automatic warning alarms, for example, if the stove is left on or if there is a water leak. [or] whether someone is within earshot before triggering interactions (Reddy et al. 2021).
		Leverage knowledge of place: [As the VUIs] is in a living room versus a bedroom, it can modify its level of proactivity, listen for particular commands, and offer suggestions for new uses (Nowacki, Gordeeva, and Lizé 2020).
Speak Users' Language (23)	Accommodate Conversational Speech (17)	Enhancing the message interactivity of the human-[VUIs] conversation by increasing the degree of contingency in message exchanges (Wang et al. 2020).
		Tailor responses and follow-up questions to make interactions more engaging, elicit in-depth disclosure, and effectively provide emotional support through these devices most natural responses (Shani et al. 2022).
	Anthropomorphize Human's Emotions (6)	Employ empathetic expressions to show emotional responsiveness. When [VUIs] use phrases such as "I understand" or "I can relate to you"... users are likely to perceive it being highly social (Wang et al. 2020).
		Express Sympathy; Be Interesting, Charming, and Lovable; Express Interest to Users (Sabir, Lafontaine, and Das 2022).
Design for Simpler Interactions (16)	Design for Shorter Conversation (8)	Design for short interactions, know when it will be long: Systems can prepare for the large majority of interactions to be a single command-answer or command-action (Axtell and Munteanu 2021).
		A11. Keep feedback and prompts short: ... the [current VUIs'] responses were not always clear or succinct, making it difficult for users to listen, understand, and remember (Wei and Landay 2018).
	Design for Simpler Conversation (8)	Minimize acoustic confusability of vocabulary (Wei and Landay 2018).
		Guide users through a conversation so they are not easily lost (Murad et al. 2019)
Establish Reliability (9)	Establish Reliability (9)	Add additional feedback currently the [VUIs] says "I will remind you" if you set a reminder, but perhaps it could also have an external visual cue that is active only if there is still a pending reminder. This additional feedback would minimize the need for participants to repeatedly ask what their reminders are and could contribute to building trust with the [VUI] system (Reddy et al. 2021).
		Acknowledgments and confirmations: To build trust, acknowledgments need to be provided as feedback indicating that the user's input was received (Alrumayh, Lehman, and Tan 2020).

References

Allworth, J. 2014. Apple's Siri is as revolutionary as the mac.

Alrumayh, A. S.; Lehman, S. M.; and Tan, C. C. 2020. Context aware access control for home voice assistant in multi-occupant homes. *Pervasive and Mobile Computing*, 67: 101196.

Axtell, B.; and Munteanu, C. 2021. Tea, Earl Grey, Hot: Designing Speech Interactions from the Imagined Ideal of Star Trek. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 1–14. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.

Bartle, V.; Lyu, J.; El Shabazz-Thompson, F.; Oh, Y.; Chen, A. A.; Chang, Y.-J.; Holstein, K.; and Dell, N. 2022. "A Second Voice": Investigating Opportunities and Challenges for Interactive Voice Assistants to Support Home Health Aides. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 1–17. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9157-3.

Begany, G.; Sa, N.; and Yuan, X.-J. 2015. Factors Affecting User Perception of a Spoken Language vs. Textual Search Interface: A Content Analysis. *Interacting with Computers*, 28: iwv029.

Blankenburg, J. 2018. Things Every Alexa Skill Should Do: Pass the One-Breath Test.

Case, A. 2015. *Calm Technology: Principles and Patterns for Non-Intrusive Design*. "O'Reilly Media, Inc.". ISBN 978-1-4919-2585-0. Google-Books-ID: DZ88CwAAQBAJ.

Chalhoub, G.; Kraemer, M. J.; Nthala, N.; and Flechais, I. 2021. "It did not give me an option to decline": A Longitudinal Analysis of the User Experience of Security and Privacy in Smart Home Products. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.

Cho, E.; Motalebi, N.; Sundar, S. S.; and Abdullah, S. 2022. Alexa as an Active Listener: How Backchanneling Can Elicit Self-Disclosure and Promote User Experience. Technical report. ArXiv:2204.10191 [cs].

Cho, M.; Lee, S.-s.; and Lee, K.-P. 2019. Once a Kind Friend is Now a Thing: Understanding How Conversational Agents at Home are Forgotten. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, DIS '19, 1557–1569. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-5850-7.

Cho, M.; and Saakes, D. 2017. Calm Automaton: A DIY Toolkit for Ambient Displays. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, 393–396. New York, NY, USA: Association for Computing Machinery. ISBN 9781450346566.

Davidoff, S.; Lee, M. K.; Yiu, C.; Zimmerman, J.; and Dey, A. 2006. Principles of Smart Home Control. volume 4206, 19–34. ISBN 978-3-540-39634-5.

Dybkjær, L.; and Bernsen, N. O. 2001. Usability evaluation in spoken language dialogue systems. In *Proceedings of the workshop on Evaluation for Language and Dialogue Systems - Volume 9*, ELDS '01, 1–10. USA: Association for Computational Linguistics.

Feng, M.; Dey, A.; and Lindeman, R. W. 2016. An initial exploration of a multi-sensory design space: Tactile support for walking in immersive virtual environments. In *2016 IEEE Symposium on 3D User Interfaces (3DUI)*, 95–104.

Fruchter, N.; and Liccardi, I. 2018. Consumer Attitudes Towards Privacy and Security in Home Assistants. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, 1–6. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356213.

Garg, R. 2022. Supporting the Design of Smart Speakers to Foster a Sense of Ownership in Asian Indian Families. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391573.

Gero, K. I.; Ashktorab, Z.; Dugan, C.; Pan, Q.; Johnson, J.; Geyer, W.; Ruiz, M.; Miller, S.; Millen, D. R.; Campbell, M.; Kumaravel, S.; and Zhang, W. 2020. Mental Models of AI Agents in a Cooperative Game Setting. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

Goetsu, S.; and Sakai, T. 2019. Voice Input Interface Failures and Frustration: Developer and User Perspectives. In *The Adjunct Publication of the 32nd Annual ACM Symposium on User Interface Software and Technology*, 24–26. New Orleans LA USA: ACM. ISBN 978-1-4503-6817-9.

Gollasch, D.; and Weber, G. 2021. Age-Related Differences in Preferences for Using Voice Assistants. In *Mensch und Computer 2021*, 156–167. Ingolstadt Germany: ACM. ISBN 978-1-4503-8645-6.

Gupta, P.; and Carew, S. 2012. Apple's Siri puts voice-enabled search in spotlight.

Jafarinaimi, N.; Forlizzi, J.; Hurst, A.; and Zimmerman, J. 2005. Breakaway: An Ambient Display Designed to Change Human Behavior. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, 1945–1948. New York, NY, USA: Association for Computing Machinery. ISBN 1595930027.

Jakob Nielsen. 1994. 10 Usability Heuristics for User Interface Design.

Javed, Y.; Sethi, S.; and Jadoun, A. 2019. Alexa's Voice Recording Behavior: A Survey of User Understanding and Awareness. In *Proceedings of the 14th International Conference on Availability, Reliability and Security*, ARES '19. New York, NY, USA: Association for Computing Machinery. ISBN 9781450371643.

Kim, S. 2021. Exploring How Older Adults Use a Smart Speaker–Based Voice Assistant in Their First Interactions: Qualitative Study. *JMIR mHealth and uHealth*, 9(1): e20427.

Kim, Y.; Reza, M.; McGrenere, J.; and Yoon, D. 2021. Designers Characterize Naturalness in Voice User Interfaces: Their Goals, Practices, and Challenges. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.

Kučera, J. S. J. 2017. Probing calmness in applications using a calm display prototype†. 5.

Lau, J.; Zimmerman, B.; and Schaub, F. 2016. "Alexa, Stop Recording": Mismatches between Smart Speaker Privacy Controls and User Needs. 6.

Lopatovska, I.; Rink, K.; Knight, I.; Raines, K.; Cosenza, K.; Williams, H.; Sorsche, P.; Hirsch, D.; Li, Q.; and Martinez, A. 2019. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science*, 51(4): 984–997. Publisher: SAGE Publications Ltd.

Luger, E.; and Sellen, A. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 5286–5297. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-3362-7.

Luo, Y.; Lee, B.; and Choe, E. K. 2020. TandemTrack: Shaping Consistent Exercise Experience by Complementing a Mobile App with a Smart Speaker. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6708-0.

Luria, M.; Hoffman, G.; and Zuckerman, O. 2017. Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, 580–628. New York, NY, USA: Association for Computing Machinery. ISBN 9781450346559.

López, G.; Quesada, L.; and Guerrero, L. A. 2018. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces. In Nunes, I. L., ed., *Advances in Human Factors and Systems Interaction*, Advances in Intelligent Systems and Computing, 241–250. Cham: Springer International Publishing. ISBN 978-3-319-60366-7.

Maguire, M. 2019. Development of a Heuristic Evaluation Tool for Voice User Interfaces. In Marcus, A.; and Wang, W., eds., *Design, User Experience, and Usability. Practice and Case Studies*, Lecture Notes in Computer Science, 212–225. Cham: Springer International Publishing. ISBN 978-3-030-23535-2.

Major, D.; Huang, D. Y.; Chetty, M.; and Feamster, N. 2021. Alexa, Who Am I Speaking To?: Understanding Users' Ability to Identify Third-Party Apps on Amazon Alexa. *ACM Transactions on Internet Technology (TOIT)*. Publisher: ACM PUB27 New York, NY.

Malkin, N.; Egelman, S.; and Wagner, D. 2019. Privacy Controls for Always-Listening Devices. In *Proceedings of the New Security Paradigms Workshop*, NSPW '19, 78–91. New York, NY, USA: Association for Computing Machinery. ISBN 9781450376471.

Matthes, J.; Karsay, K.; Schmuck, D.; and Stevic, A. 2020. "Too much to handle": Impact of mobile social networking sites on information overload, depressive symptoms, and well-being. *Computers in Human Behavior*, 105: 106217.

Mavrina, L.; Szczuka, J.; Strathmann, C.; Bohnenkamp, L. M.; Krämer, N.; and Kopp, S. 2022. "Alexa, You're Really Stupid": A Longitudinal Field Study on Communication Breakdowns Between Family Members and a Voice Assistant. *Frontiers in Computer Science*, 4: 791704.

Murad, C.; Munteanu, C.; Clark, L.; and Cowan, B. R. 2018. Design Guidelines for Hands-Free Speech Interaction. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, MobileHCI '18, 269–276. New York, NY, USA: Association for Computing Machinery. ISBN 9781450359412.

Murad, C.; Munteanu, C.; Cowan, B. R.; and Clark, L. 2019. Revolution or Evolution? Speech Interaction and HCI Design Guidelines. *IEEE Pervasive Computing*, 18(2): 33–45. Conference Name: IEEE Pervasive Computing.

Murad, C.; Munteanu, C.; R. Cowan, B.; and Clark, L. 2021. Finding a New Voice: Transitioning Designers from GUI to VUI Design. In *CUI 2021 - 3rd Conference on Conversational User Interfaces*, 1–12. Bilbao (online) Spain: ACM. ISBN 978-1-4503-8998-3.

National Public Radio. 2020. Use of Smart Speakers in the U.S. Increases During Quarantine. *NPR*.

National Public Radio. 2022. NPR & Edison Research: Smart Speaker Ownership Reaches 35% of Americans. *NPR*.

Nowacki, C.; Gordeeva, A.; and Lizé, A.-H. 2020. Improving the Usability of Voice User Interfaces: A New Set of Ergonomic Criteria. In Marcus, A.; and Rosenzweig, E., eds., *Design, User Experience, and Usability. Design for Contemporary Interactive Environments*, Lecture Notes in Computer Science, 117–133. Cham: Springer International Publishing. ISBN 978-3-030-49760-6.

Occhialini, V.; van Essen, H.; and Eggen, B. 2011. Design and evaluation of an ambient display to support time management during meetings. *Human-Computer Interaction – INTERACT 2011*, 263–280.

Olmstead, K. 2017. Nearly half of Americans use digital voice assistants, mostly on their smartphones.

Pan, L.; Yu, C.; Li, J.; Huang, T.; Bi, X.; and Shi, Y. 2022. Automatically Generating and Improving Voice Command Interface from Operation Sequences on Smartphones. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. New York, NY, USA: Association for Computing Machinery. ISBN 9781450391573.

Pearl, C. 2017. Designing Voice User Interfaces. 278.

Pielot, M.; and Rello, L. 2015. The Do Not Disturb Challenge: A Day Without Notifications. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '15, 1761–1766. New York, NY, USA: Association for Computing Machinery. ISBN 9781450331463.

Purington, A.; Taft, J. G.; Sannon, S.; Bazarova, N. N.; and Taylor, S. H. 2017. "Alexa is my new BFF": Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, CHI EA '17, 2853–2859. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-4656-6.

Ramadan, Z.; Farah, M.; and El Essrawi, L. 2021. From Amazon.com to Amazon.love: How Alexa is redefining companionship and interdependence for people with special needs. *Psychology & Marketing*, 38(4): 596–609.

Reddy, A.; Kocaballi, A. B.; Nicenboim, I.; Søndergaard, M. L. J.; Lupetti, M. L.; Key, C.; Speed, C.; Lockton, D.; Giaccardi, E.; Grommé, F.; Robbins, H.; Primlani, N.; Yurman, P.; Sumartojo, S.; Phan, T.; Bedö, V.; and Strengers, Y. 2021. Making Everyday Things Talk: Speculative Conversations into the Future of Voice Interfaces at Home. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI EA '21, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8095-9.

Reeves, S.; Porcheron, M.; Fischer, J. E.; Candello, H.; McMillan, D.; McGregor, M.; Moore, R. J.; Sikveland, R.; Taylor, A. S.; Velkovska, J.; and Zouinar, M. 2018. Voice-Based Conversational UX Studies and Design. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI EA '18, 1–8. New York, NY, USA: Association for Computing Machinery. ISBN 9781450356213.

Reportlinker. 2022. Smart Speakers Global Market Report 2022.

Sabir, A.; Lafontaine, E.; and Das, A. 2022. Hey Alexa, Who Am I Talking to?: Analyzing Users' Perception and Awareness Regarding Third-party Alexa Skills. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 1–15. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9157-3.

Sciuto, A.; Saini, A.; Forlizzi, J.; and Hong, J. I. 2018. "Hey Alexa, What's Up?": A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*, DIS '18, 857–868. New York, NY, USA: Association for Computing Machinery. ISBN 9781450351980.

Setlur, V.; and Tory, M. 2022. How do you Converse with an Analytical Chatbot? Revisiting Gricean Maxims for Designing Analytical Conversational Behavior. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 1–17. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9157-3.

Shani, C.; Libov, A.; Tolmach, S.; Lewin-Eytan, L.; Maarek, Y.; and Shahaf, D. 2022. "Alexa, Do You Want to Build a Snowman?" Characterizing Playful Requests to Conversational Agents. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI EA '22, 1–7. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9156-6.

Shedroff, N. 2000. 11 Information Interaction Design: A Unified Field Theory of Design. *Information design*, 267.

Striegl, J.; Gollasch, D.; Loitsch, C.; and Weber, G. 2021. Designing VUIs for Social Assistance Robots for People with Dementia. In *Mensch Und Computer 2021*, MuC '21, 145–155. New York, NY, USA: Association for Computing Machinery. ISBN 9781450386456.

Tabassum, M.; Kosiński, T.; Frik, A.; Malkin, N.; Wijesekera, P.; Egelman, S.; and Lipford, H. R. 2019. Investigating Users' Preferences and Expectations for Always-Listening Voice Assistants. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 3(4).

Team, S. 2017. Deep learning for siri's voice: On-device deep mixture density networks for hybrid unit selection synthesis.

Tentori, M.; Segura, D.; and Favela, J. 2009. Monitoring hospital patients using ambient displays. *Mobile Health Solutions for Biomedical Applications*, 143–158.

Trajkova, M.; and Martin-Hammond, A. 2020. "Alexa is a Toy": Exploring Older Adults' Reasons for Using, Limiting, and Abandoning Echo. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 9781450367080.

Vacher, M.; Caffiau, S.; Portet, F.; Meillon, B.; Roux, C.; Elias, E.; Lecouteux, B.; and Chahuara, P. 2015. Evaluation of a Context-Aware Voice Interface for Ambient Assisted Living: Qualitative User Study vs. Quantitative System Evaluation. *ACM Transactions on Accessible Computing*, 7(2): 1–36.

Voit, A.; Niess, J.; Eckerth, C.; Ernst, M.; Weingärtner, H.; and Woźniak, P. W. 2020. 'It's Not a Romantic Relationship': Stories of Adoption and Abandonment of Smart Speakers at Home. In *19th International Conference on Mobile and Ubiquitous Multimedia*, MUM '20, 71–82. New York, NY, USA: Association for Computing Machinery. ISBN 9781450388702.

Wang, J.; Yang, H.; Shao, R.; Abdullah, S.; and Sundar, S. S. 2020. Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 1–13. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-6708-0.

Wei, Z.; and Landay, J. A. 2018. Evaluating Speech-Based Smart Devices Using New Usability Heuristics. *IEEE Pervasive Computing*, 17(2): 84–96.

Weiser, M. 1991. The Computer for the 21st Century. 8.

Weiser, M. 1998. The Future of Ubiquitous Computing on Campus. *Commun. ACM*, 41(1): 41–42.

Weiser, M.; and Brown, J. S. 1995. Designing Calm Technology.

Weiser, M.; and Brown, J. S. 1996. The Coming Age of Calm Technology.

Yang, X.; Aurisicchio, M.; and Baxter, W. 2019. Understanding Affective Experiences with Conversational Agents.

In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, 1–12. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-5970-2.

Yu, C.; Shane, H.; Schlosser, R. W.; O'Brien, A.; Allen, A.; Abramson, J.; and Flynn, S. 2018. An exploratory study of speech-language pathologists using the Echo Show™ to deliver visual supports. *Advances in Neurodevelopmental Disorders*, 2(3): 286–292.

Zhang, X.; Su, Z.; and Rekimoto, J. 2022. Aware: Intuitive Device Activation Using Prosody for Natural Voice Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 1–16. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9157-3.

Zhang, Z.; Xu, Y.; Wang, Y.; Yao, B.; Ritchie, D.; Wu, T.; Yu, M.; Wang, D.; and Li, T. J.-J. 2022. StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, 1–21. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-9157-3.

Zubatiy, T.; Vickers, K. L.; Mathur, N.; and Mynatt, E. D. 2021. Empowering Dyads of Older Adults With Mild Cognitive Impairment And Their Care Partners Using Conversational Agents. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, 1–15. New York, NY, USA: Association for Computing Machinery. ISBN 978-1-4503-8096-6.

Zwakman, D. S.; Pal, D.; and Arpnikanondt, C. 2021. Usability Evaluation of Artificial Intelligence-Based Voice Assistants: The Case of Amazon Alexa. *Sn Computer Science*, 2(1): 28.